



Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Stereo depth estimation using synchronous optimization with segment based regularization

Tarkan Aydin^{a,b,*}, Yusuf Sinan Akgul^a^a GIT Vision Lab, Department of Computer Engineering, Gebze Institute of Technology, Gebze, Kocaeli 41400, Turkey^b Bahcesehir University, Istanbul 34353, Turkey

ARTICLE INFO

Article history:

Received 28 January 2010

Available online xxxxx

Communicated by A. Fernandez-Caballero

Keywords:

Stereo

Optimization

Segment based regularization

Anisotropic smoothing

ABSTRACT

Stereo correspondence is inherently an ill-posed problem, which is addressed by regularization methods. This paper introduces a novel stereo correspondence method that uses two synchronous interdependent optimizations. The regularization of the correspondence problem is done adaptively by considering the image segments and the intermediate disparity maps of the two optimizations. Our adaptive regularization allows inter-segment diffusion at the beginning of the optimizations to be robust against local minima. When the two optimizations start producing similar disparity maps, our regularization prevents inter-segment diffusion to recover the depth discontinuities. Our experimental results showed that the proposed algorithm can handle sharp discontinuities well and provides disparity maps with accuracy comparable to the state of the art stereo methods.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Stereo correspondence is one of the fundamental problems of computer vision. The typical result of the correspondence problem is expressed as a disparity map, i.e. spatial shifts in the pixel positions of the corresponding points. The main difficulties of the correspondence problem are the ambiguity due to the image noise, repeated texture, and occlusions. These problems make the stereo correspondence an ill-posed problem, which is classically addressed by a regularization method to stabilize the solution. The main role of the regularization is to incorporate a priori information to handle image noise and to fill-in missing and ambiguous data.

Classically, the regularization is employed by local and global methods. The local methods perform regularization directly on the data space by employing some aggregation scheme (Intille and Bobick, 1994; Kanade and Okutomi, 1994; Scharstein and Szeliski, 1998; Yoon and Kweon, 2006). The global methods, on the other hand, formulate the problem as an energy functional that needs to be minimized to produce the desired solution. The regularization is performed on the disparity space by introducing an explicit smoothness criteria so that reliable disparity values are

propagated to ambiguous image regions. Dynamic programming was tried by enforcing smoothness only along the epipolar lines in order to obtain a globally optimal solution to the discrete form of the energy functional (Ohta and Kanade, 1985; Gong and Yang, 2005). However, the resulting disparity maps contain well-known streaking effects due to the inconsistency between epipolar lines. Alternatively, a global minimum of the functional can also be obtained in polynomial time via graph cuts (Roy and Cox, 1998; Ishikawa, 2003) by using a convex smoothness term. However, these methods oversmooth the depth discontinuities. A discontinuity preserving regularizer might produce a good solution but it is known that introducing a discontinuity preserving smoothness term makes the problem NP-complete (Kolmogorov and Zabih, 2004). Therefore, using an approximate optimization method for the functional with non-convex smoothness terms became more popular, such as graph cuts (Boykov et al., 2001), belief propagation (Sun et al., 2003), and genetic algorithm (Saito and Mori, 1995). However, these methods only produce integer valued disparity maps due to their discrete nature. This restriction is a severe drawback if curved or slanted surfaces are present in the scene (Li and Zucker, 2006).

Another class of global approaches, as a counterpart, use related partial differential equations (PDE) and variational methods in order to find the minimizer of the continuous form of the energy functional. These methods can achieve a continuous solution by iteratively evaluating the associated Euler–Lagrange equation. An inherent advantage of these methods is the capability of making sub-pixel disparity estimations due to the continuous solution they provide.

* Corresponding author at: GIT Vision Lab, Department of Computer Engineering, Gebze Institute of Technology, Gebze, Kocaeli 41400, Turkey. Tel.: +90 262 6052243; fax: +90 262 6052205.

E-mail addresses: tarkan.aydin@bahcesehir.edu.tr (T. Aydin), akgul@bilmuh.gyte.edu.tr (Y.S. Akgul).

URL: <http://vision.gyte.edu.tr> (T. Aydin).

The minimization process of these continuous methods is characterized by the choice of the regularizer. Using a disparity driven isotropic regularizer with a quadratic term makes the minimization robust against local minima (Robert et al., 1992). However, the depth discontinuities in the resulting disparity maps would be oversmoothed. Although it is possible to use a non-quadratic smoothing term, such as total variation regularizer, to inhibit the oversmoothing of discontinuities (Slesareva et al., 2005), it cannot handle the discontinuities adequately (Ben-Ari and Sochen, 2007).

There are several other regularization methods for the handling of the discontinuities. Shah (1993) uses nonlinear diffusion to extract stereo matches and occluded regions simultaneously in conjunction with a gradient descent minimization. Similarly, Robert and Deriche (1996) use anisotropic disparity driven regularization in order to prevent smoothing of the disparity map at the estimated discontinuities. Therefore, it tends to preserve the discontinuities present at the initialization. Alternatively, image driven regularizers were used to align depth discontinuities along edges and inhibit smoothing across edges (Alvarez et al., 2002; Kim et al., 2004). Min et al. (2006) employed the image segments to perform anisotropic smoothing at the segment boundaries depending on the magnitude of image gradients. The problem with image driven regularizers is that they have to work with over-segmented images when the images are highly textured. In addition, the boundary leakage problem becomes an issue when there are gaps at the object boundaries.

Nevertheless, these discontinuity preserving approaches require sufficiently reliable initialization in order to converge to the desired solution. In most cases, the initialization errors cannot be recovered, especially for noisy and occluded regions.

In this paper, we introduce a novel initialization insensitive regularization method that preserves the depth discontinuities. Our framework employs two separate but dependent energy functionals (Akgul and Kambhamettu, 1999; Aydin and Akgul, 2006) which are intended to be minimized synchronously until converging to the same solution. Because of the interaction between the optimizations, the overall result of our system is always better than the results achievable by a single optimization. Reliable convergence is ensured by starting each optimization with different initial conditions.

In order to handle depth discontinuities robustly, we employ image segments to align the depth discontinuities with the segment boundaries. Unlike the previous image based smoothing techniques, the proposed method adjusts the smoothing by utilizing not only the segment information but also the positional differences between the synchronous optimizations. These two means of adjusting the smoothing make it possible to use isotropic and anisotropic smoothing adaptively. As a result, we produce more robust depth discontinuity positions. Note that our employment of synchronous optimizations is very different from that of Aydin and Akgul (2006), which does not use the optimizations for the regularization and completely ignores the depth discontinuities.

Selecting an appropriate stopping criteria is crucial for many diffusion techniques in order to avoid oversmoothing and insufficient regularization (Scharstein and Szeliski, 1998). Since our discontinuity preserving regularization method relies on the positional difference between the solutions of each optimization, the diffusion between the segments are prevented when both optimizations find the same disparity map, hence smoothing of the discontinuities is inhibited even at superfluous iterations. This inherent stopping criteria of our framework is an important advantage over the similar systems against problems like sensitivity to extra iterations.

The rest of this paper is organized as follows. Section 2 reviews the synchronous energy functional. Section 3 describes the proposed regularization that preserves the depth discontinuities. Sec-

tion 4 describes the system validation and experiments. Finally, we provide concluding remarks in Section 5.

2. Overview of the approach

2.1. Energy-based global stereo formulation

Traditional global stereo energy formulation is written as the sum of the data term and a regularization term. Consequently, the stereo correspondence problem is formulated as the minimization of the following energy functional,

$$E(D) = \int \alpha\phi(D) + \beta\psi(|\nabla D|)dp, \quad (1)$$

where D is the disparity map which assigns disparity values to each pixel p in the reference image. α and β are weighting coefficients for adjusting the relative weights of each term.

The data term ϕ computes the image similarity measure by means of commonly used similarity metrics, such as the sum of squared differences (SSD), the sum of absolute differences (SAD), and the normalized cross correlation (NCC). The smoothness or regularization term ψ is introduced to impose a priori information (smoothness) on the desired disparity map by penalizing disparity gradients (∇D).

2.2. Synchronous energy formulation

Based on the classical stereo energy functional, the synchronous optimizations are formulated as the minimization of two energy functionals by introducing a new tension term φ as in the following equations.

$$E(D_1) = \int \alpha\phi(D_1) + \beta\psi(|\nabla D_1|^2) + \lambda\varphi((D_1 - D_2)^2)dp, \quad (2)$$

$$E(D_2) = \int \alpha\phi(D_2) + \beta\psi(|\nabla D_2|^2) + \lambda\varphi((D_2 - D_1)^2)dp, \quad (3)$$

where D_1 and D_2 are the disparity maps obtained from each optimization.

The tension term φ is for the interaction between the two minimizations and it is the core idea of the synchronous optimization method. The main function of this term is to lower the difference between the two disparity maps D_1 and D_2 . Note that without the tension term, minimization of the energy functionals defined by Eqs. (2) and (3) by starting from different initial configurations would produce a different disparity map for each equation. However, if the equations are optimized in synchronization with the help of the tension term, they would end up finding the same disparity map.

The disparity maps are computed by searching the minimizers of the energy functionals defined in Eqs. (2) and (3). Minimization of the functionals via the gradient descent method by introducing an artificial evolution parameter t yields the equations,

$$\frac{\partial D_1}{\partial t} = \gamma(\alpha\phi'(D_1) + \beta\nabla \cdot (\psi'\nabla D_1) + \lambda\varphi'(D_1 - D_2)), \quad (4)$$

$$\frac{\partial D_2}{\partial t} = \gamma(\alpha\phi'(D_2) + \beta\nabla \cdot (\psi'\nabla D_2) + \lambda\varphi'(D_2 - D_1)), \quad (5)$$

where ϕ' , ψ' and φ' are the derivatives of the functions ϕ , ψ and φ , respectively. ψ' is also called diffusion or conduction coefficient (Perona and Malik, 1990). The minimizers are found by computing asymptotic states ($t \rightarrow \infty$) of the solutions D_1^t and D_2^t , which are the disparity maps produced by first and second optimizations at iteration t .

The function of the diffusion term is to produce disparity maps that assign similar values to neighboring pixels if there is no depth discontinuity between the pixels, which is called regularization.

Similarly, the function of the tension term is to make the two synchronous optimizations produce similar disparity maps, which can be considered as another form of regularization. This extra regularization helps further our system address the inherent ill-posedness of the stereo correspondence problem.

The tension term forces both optimizations to converge to the same solution. However, continually forcing both optimizations towards each other may result in convergence to an irrelevant local minimum. In order to address this problem, the tension coefficient φ' is constructed such that the optimization with the smaller data term values is not affected from this term. The coefficient for the first optimization is defined as

$$\varphi' = \begin{cases} 1 - e^{\left(\frac{\Delta\phi}{\kappa_\varphi}\right)^2} & \Delta\phi \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where $\Delta\phi = \phi(D_1^t) - \phi(D_2^t)$, and κ_φ is a constant. Similarly, the same coefficient is computed for the second optimization.

Note that the tension term is not symmetric and it depends on the intermediate disparity maps of the optimizations. As a result, it computes a different value for each optimization. If the optimization has larger data term values than the other, it will be pulled by the tension term towards the other optimization. This procedure eliminates a considerable amount of local minima problems because when one optimization is stuck due to local minima, it is always possible to compare the position with the other optimization. As a result, the overall optimization would localize a better position than each of the optimizations can achieve by themselves.

In order to find the desired minimum, initialization steps are very critical for the classical iterative optimization methods. Initial positions far away from the global minimum results in converging to a local minimum in most cases. However, in our system, initialization of the optimizations does not depend on any prior assumption about the global minimum and it is always done the same way. One optimization is initialized with the minimum disparity (d_{min}) values whereas the other is initialized with the maximum disparity (d_{max}) values.

$$\begin{aligned} D_1^0 &= d_{min}, \\ D_2^0 &= d_{max}. \end{aligned} \quad (7)$$

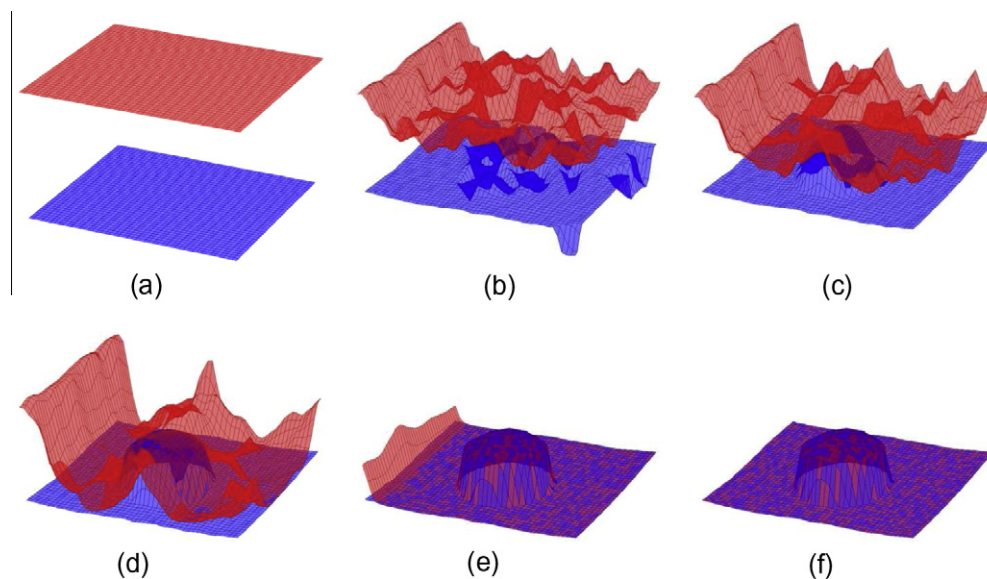


Fig. 1. Sample intermediate disparity maps of synchronous optimizations for the stereo pair shown in Fig. 2. (a) Initialization step of the optimizations are always the same. (b–e) Disparity maps D_1^t and D_2^t evolve to find the final disparity map. (f) Discontinuities are recovered accurately at the end of the optimizations.

A sample optimization can be seen in Fig. 1. As the figure shows, one optimization starts from the minimum disparity values and the second optimization starts from the maximum disparity values (Fig. 1a). During the system iterations, disparity values of the optimizations get close to each other (Fig. 1b–e). The optimizations continue until both of them find the same disparity map (Fig. 1f) and the depth discontinuities are also recovered. The final disparity map of the tennis ball is shown in Fig. 2. As seen in Fig. 2c, the depth discontinuities are also recovered robustly by means of our novel regularizer. Fig. 2d shows the final disparity map obtained by Aydin and Akgul (2006), which does not preserve discontinuities.

The next section describes our novel regularizer that employs the image segment information along with the intermediate disparity maps of optimizations.

3. Synchronous optimization with segment based regularization

The classical regularization is used to solve an ill-posed problem to reconstruct a well-posed form by adding a stabilizing term to the initial formulation. The stabilizing term incorporates a priori information into the formulation in the form of a constraint, such as smoothness constraint, that will be imposed on the solution.

The most common and well-known type of regularization is *Tikhonov* regularization which quadratically penalizes the gradient of the solution and inherently implies a globally smooth solution (Tikhonov and Arsenin, 1977). Since this approach ignores the discontinuities, its direct application for the stereo correspondence problem leads to the blurring of the disparity map across the depth discontinuities. Hence, only an approximate coarse structure of the actual surface can be recovered. A proper regularizer should take the discontinuities into consideration in order to recover disparity maps accurately.

One important design issue for a discontinuity preserving regularizer is the selection of a discontinuity marker function. The discontinuity marker function decides the presence and the location of the discontinuities so that smoothing of them can be prevented. If the locations of the discontinuities were known in advance, they could have been easily incorporated into the regularization. However, this information is not available, and it is actually an essential

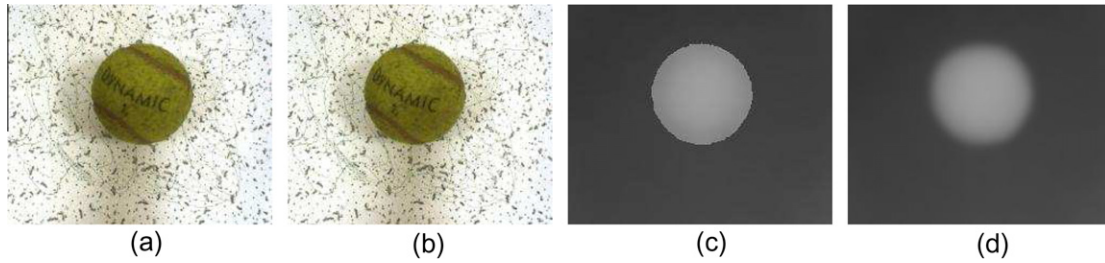


Fig. 2. (a) Left and (b) right images of tennis ball stereo pairs. (c) Computed disparity map using our method and (d) synchronous processes (Aydin and Akgul, 2006).

part of the solution that we seek. Therefore, it should be estimated or inferred from a reasonable source.

Disparity-driven (or solution-driven) regularizers infer the locations of the discontinuities from the intermediate solutions by assuming that a discontinuity in the data term implies a depth discontinuity (Robert and Deriche, 1996). However, it is very hard to determine whether the emergent discontinuities are caused by the depth discontinuities or ambiguity in data term due to occlusion, repeated pattern, noise, etc.

Image driven regularizers, on the other hand, utilize the intensity discontinuities in the images to estimate locations of depth discontinuities (Alvarez et al., 2002; Kim et al., 2004). Therefore, they generally offer better performance than the disparity driven regularizers near the image edges. These regularizers work under the assumption that depth discontinuities coincide with some intensity discontinuities in the image. Thus, they allow isotropic smoothing in homogeneous regions and prevent smoothing in inhomogeneous regions by simply adjusting the degree of smoothing according to the magnitude of image gradients.

However, using image gradient magnitude information to determine the homogenous region boundaries does not produce consistent results because region boundaries cannot be estimated reliably using the image gradient magnitude values which are too local. Using this local information to adjust the degree of smoothing results in the boundary leakage problem, which is the leakage of diffusion across the region boundaries (see Fig. 3b and c). In addition, on some noisy image regions, diffusion anticipated between the elements of the homogeneous regions is prevented. Finally, this approach might apply different degrees of smoothing for the regions with the same depth discontinuity values because the magnitudes of discontinuities are not directly correlated with the magnitudes of the image gradients.

Instead of basing the depth discontinuity decisions on local image gradient magnitudes, we use image segment boundaries as the indicators of depth discontinuities. There are sophisticated image segmentation methods that utilize information about the intra-segment homogeneity and inter-segment inhomogeneity. Therefore, more global information would be incorporated into the esti-

mation of the depth discontinuity positions if we employ segment information from such methods. The relatively global nature of the depth discontinuity decisions makes the overall system robust against the problems of image driven regularization methods.

In segment based regularizers, the diffusion between the pixels of the same segment is always allowed because it is assumed that there are no depth discontinuities inside the segments. The diffusion between the neighboring segments, however, should be handled very carefully. The complete prevention of the diffusion between the segments would result in a system with very serious local minima problems because each segment has to behave independently. The other extreme of allowing full diffusion between the segments would not allow any depth discontinuity localizations. As a result, there has to be an efficient method for deciding the amount of diffusion between segments. We take advantage of our system of synchronous optimizations to effectively address this problem by utilizing the difference between the intermediate disparity maps of optimizations. The difference Δd between the intermediate disparity maps of the optimizations at iteration t is calculated by

$$\Delta d(x, y, t) = D_1^t(x, y) - D_2^t(x, y). \quad (8)$$

One of the synchronous optimizations is started from minimum disparity values and the other is started from maximum disparity values. Initially Δd has the maximum possible value. At this time, the regularization should be isotropic to avoid getting stuck to local minima. In other words, unconstrained smoothing is allowed between the neighboring segments. During the minimizations, the disparity maps of each optimization get close to each other and Δd becomes smaller. In order to prevent smoothing of discontinuities, the regularization should behave anisotropically as the optimizations approach the desired solution. Eventually, Δd becomes zero and no diffusion between the segments are allowed for a full recovery of discontinuities.

We include Δd in the diffusion function ψ' in Eqs. (4) and (5) to adaptively adjust the degree of smoothing based on both segment boundaries and difference between the intermediate disparity maps. The diffusion function ψ' is defined as,

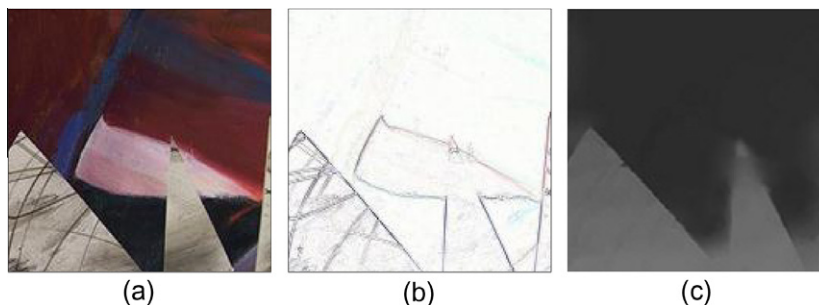


Fig. 3. An example illustrates the boundary leakage problem. (a) Clipped part of the Sawtooth image, (b) gradient magnitudes of the image, and (c) resulting disparity map obtained with an image driven regularizer (Kim et al., 2004). Note that oversmoothing occurs at object boundaries.

$$\psi' = g(\nabla I_s(x, y), \Delta d(x, y)). \quad (9)$$

The function g is known as the discontinuity marker function and it is defined as,

$$g(\nabla I_s, \Delta d) = 1 - \nabla I_s e^{-(\Delta d / \kappa_\psi)^2}, \quad (10)$$

where I_s is the segmented image and κ_ψ is the system parameter that controls the adaptivity of the regularizer. The gradient of the segmented image is defined as,

$$\nabla I_s(x, y) = \begin{cases} 1, & \text{if } (x, y) \text{ is at the segment boundary,} \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

If Δd is large, the diffusion function evaluates to one. It means that smoothing is isotropic which is the case in *Tikhonov* regularizer. Consequently, the minimization is not affected by local minima. When the optimizations get close to each other, the smoothing gradually becomes anisotropic as in the image driven regularizers. Eventually, the optimizations find the same disparity maps ($\Delta d = 0$) due to the tension term and the diffusion function evaluates to zero at segment boundaries. At this time, the regularizer exhibits pure anisotropic behavior. This is what is expected from a discontinuity preserving regularizer: smoothing is enforced only in homogenous regions. An additional advantage of this approach is that no oversmoothing artifacts can be introduced into the system even if superfluous iterations are executed after optimum solution is achieved. The alternative regularizers, such as image driven or disparity driven regularizers, there is always the risk of oversmoothing of discontinuities because there is always diffusion between the neighboring elements.

4. Experimental results

The proposed method is tested on Middlebury (Scharstein and Szeliski, 2003) data sets where the ground truth information is available for benchmarking. The image segments used in the diffusion function are obtained from the left image of the stereo pairs by applying the mean shift segmentation algorithm (Comaniciu and Meer, 2002).

We use normalized cross correlation method as the similarity measure due to its robustness against any brightness differences. In order to increase the performance further and decrease the convergence time of our method, data space should be pre-smoothed. Employing higher window sizes in the evaluation of correlation values basically satisfies this condition. However, increasing the correlation window sizes results in shifts at the location of discontinuities (Scharstein and Szeliski, 2002). The discontinuities must be preserved in the data space in order to recover them accurately. Therefore, we pre-smooth data space with a bilateral filter (Tomasi and Manduchi, 1998) whose kernels are derived from the left image of stereo pairs. This smoothing strategy is similar to the method of Yoon and Kweon (2006), in which they employ the product of

two bilateral filters which are derived from the left and the right stereo images.

Experiments are performed on Venus, Teddy, Tsukaba, and Cones data sets that contain sharp depth discontinuities. The system parameters kept fixed during the experiments ($\alpha = 0.1$, $\beta = 0.1$, $\lambda = 0.15$, $\gamma = 0.1$, and $\kappa_\psi = 0.01$). The parameter of diffusion function is associated with the maximum disparity range in the setup ($\kappa_\psi = d_{max} - d_{min}$) which is different for each data set. Smoothing of data space is performed with 11×11 pixels sized bilateral filter. Error rates of the proposed algorithm are computed for non-occluded areas, near discontinuities, and for complete images. Table 1 compares the error rates of our results with the results of the state of the art segment based global methods and local methods employing pre-smoothed data values. Error rates in the table are calculated by setting the error threshold values to one pixel disparity. Fig. 4 shows the left stereo images, the segment boundaries used in our adaptive regularizer and the resulting disparity maps of our algorithm. The visual and numerical results show that our method can robustly recover piecewise smooth surfaces and preserve discontinuities well. Since our method works on continuous disparity space, we compute the error rates by setting the threshold value to 0.5 pixels in order to show the performance of our method also in sub-pixel disparity estimation. The sub-pixel error rates of our method and some other methods are shown in Table 2. The analysis of these numbers indicates that, there is no clear best correspondence method for all types of images. Furthermore, the methods show different performance rates when the error threshold is changed pixel level to sub-pixel. The experiments also show that our proposed method produces similar error rates as the other leading methods. It should be noted that some of these methods employ extra information from the images such as plane fitting, multiple image segmentation or they enhance their results using subsequent disparity processing methods.

As the final experiment, we produced data to show the insensitivity of our method against extra optimization iterations. Table 3 shows the changes in the error rates with respect to the number of iterations. It can be easily seen that recovered discontinuities are not smoothed and the error rates do not change even if optimizations are forced to continue for a longer period of time.

We observed that the method fails mostly near the image sections that include very small or problematic image segments that are difficult to segment. Our method depends on the results of the image segmentation and if the segmentation is not correct, it does not have any ways of correcting them. If the depth discontinuities do not coincide with the segment boundaries, our method might not produce the correct depth values. This problem can be addressed by employing multiple segmented images (Woodford et al., 2008). The second source of error is near the image sections that do not include sufficient image texture. Although our regularization method can handle considerable amount of textureless regions due to the employment of multiple

Table 1

The disparity error percentages (disparity error threshold of 1 pixel) using the Middlebury stereo benchmark data sets. Some results in the table are missing because they are not available in the original publication.

Algorithm	Tsukaba			Venus			Teddy			Cones		
	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc
Woodford et al. (2008)	2.91	3.56	0.24	0.24	0.49	2.76	10.9	15.4	20.6	5.42	10.8	12.5
Yang et al. (2007)	1.24	1.76	5.98	0.12	0.46	1.74	3.45	8.38	10.0	2.93	8.73	7.91
Yoon and Kweon (2006)	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26
Ben-Ari and Sochen (2010)	3.97	5.23	14.9	0.28	0.76	3.78	9.34	14.3	20.0	4.14	9.91	11.4
Pock et al. (2008)	3.61	5.72	18.0	1.16	2.50	12.4	6.10	15.7	16.8	3.88	14.4	11.5
Pock et al. (2007)	–	–	–	–	1.10	–	–	6.63	–	–	3.67	–
Our method	2.51	2.71	7.13	0.32	0.40	3.45	8.08	12.3	21.2	4.08	11.5	13.7

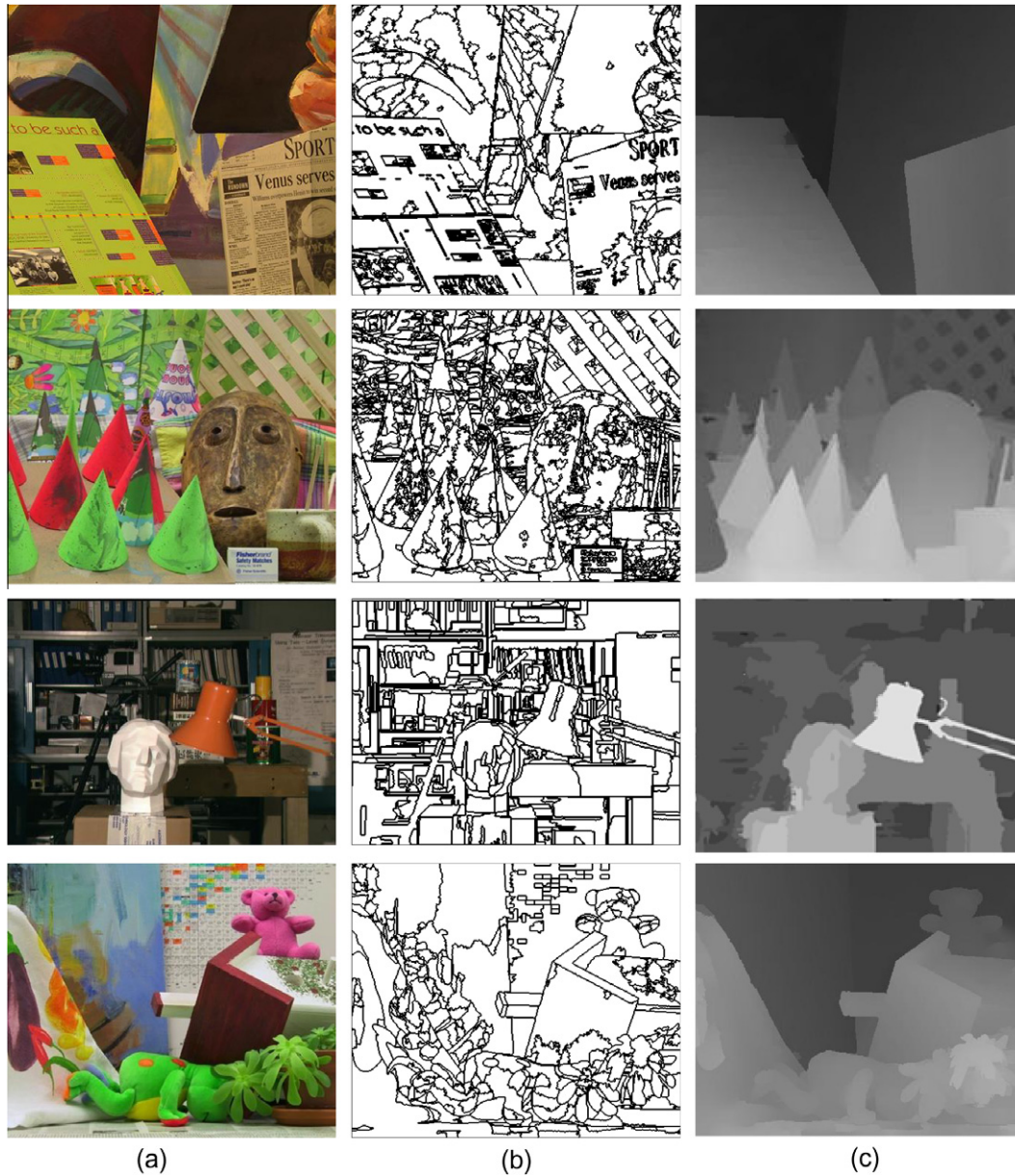


Fig. 4. (a) The data sets (Venus, Cones, Tsukaba, and Teddy sets) from Middlebury. (b) Boundaries of segments used in our adaptive regularization. (c) Computed disparity maps.

Table 2
The disparity error percentages (disparity error threshold of 0.5 pixel) using the Middlebury stereo benchmark data sets. Some results in the table are missing because they are not available in the original publication.

Algorithm	Tsukaba			Venus			Teddy			Cones		
	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc
Woodford et al. (2008)	7.10	7.70	0.56	0.56	0.83	4.21	17.5	22.7	31.5	11.6	17.1	20.4
Yang et al. (2007)	8.78	9.45	14.9	0.72	1.12	5.24	10.1	16.4	21.3	8.49	14.7	16.5
Yoon and Kweon (2006)	18.1	18.8	18.6	7.77	8.40	15.83	17.6	23.9	34.0	14.0	19.7	20.6
Ben-Ari and Sochen (2010)	7.18	8.56	20.1	1.46	2.12	7.87	12.9	19.4	27.5	6.22	12.6	15.8
Pock et al. (2008)	11.1	13.3	27.2	5.99	7.40	22.3	10.5	19.9	25.8	5.99	16.5	16.7
Pock et al. (2007)	–	–	–	–	3.45	–	–	11.2	–	–	7.52	–
Our method	8.53	9.29	16.98	1.23	1.52	9.55	10.2	16.1	25.4	7.34	14.2	14.4

processes, large areas of insufficient texture still causes problems. Finally, our method does not include any explicit occlusion detec-

tion mechanisms and occluded regions could produce high error rates.

Table 3

The disparity errors generated for the Venus image with respect to the number of iterations.

Iteration	Disparity error 1 pixel			Disparity error 0.5 pixel		
	Nonocc	All	Disc	Nonocc	All	Disc
1000	27.6	28.3	35.4	31.0	31.9	47.0
1400	12.4	12.9	26.0	16.6	17.2	33.2
1800	1.91	2.29	12.0	3.17	3.77	19.6
2000	0.95	1.35	9.38	1.61	2.04	10.9
2500	0.32	0.40	3.45	1.23	1.52	9.55
10,000	0.32	0.40	3.45	1.23	1.52	9.55

5. Conclusions

Stereo correspondence problem is an inverse problem and like most inverse problems, its mathematical formulation is ill-posed. The most common approach for solving these types of problems is to regularize the solution by including additional information or making prior assumptions about the solution.

In this paper, we proposed a novel system that uses two energy functionals which are minimized synchronously by two dependent optimizations. The new regularizer can recover piecewise smooth disparity maps from stereo image pairs without blurring the discontinuities. It employs segmented version of the image and the positional differences between the optimizations. Including these additional information in regularization results in an adaptive smoothing around the segment boundaries by adjusting the degree of smoothing depending on the intermediate values of the optimizations. Consequently, initial isotropic smoothing gradually turns into an anisotropic smoothing.

The system addresses many problems common to approximate optimization methods, such as sensitivity to initializations and local minima. Therefore, the final recovered disparity surface turns out to be more accurate than what a single optimization can achieve.

The system also addresses the problem of selecting optimal stopping criteria which is a very important step for the diffusion based methods to avoid smoothing of the discontinuities. When the optimizations find the same solution, the diffusion between the segments is prevented. Consequently, the system does not suffer from superfluous iterations executed after the optimum solution is achieved.

Experiments performed with the standard Middlebury stereo pairs show the accuracy of our method around the homogenous and nonhomogeneous image regions. The results are found to be comparable to the state of the art stereo methods.

The current limitations of the system include the handling of the occluded image regions as regular image regions. Segments located completely inside the occluded regions cannot be recovered in most cases. Nevertheless, due to the new regularization strategy introduced, our system can produce acceptable results on these regions but an explicit occlusion mechanism would make our system much more robust.

Another limitation of the method is higher computational complexity due to two dependent optimizations. Our current running time is around 20 min for Venus image (434 × 383 pixels) on a 1.66 GHz PC with 1 GB RAM. Although the convergence time of the method can be decreased by adapting a multi-scale approach, the adaptation is not trivial for our method. A sample implementation of multi-scale approach can be found in the work of Akgul and Kambhamettu (1999). Another speed up path could be an implementation on a GPU architecture because it is known that PDE-based approaches are generally suitable for parallel computing.

Acknowledgements

This work was conducted at the Computer Vision Laboratory at Gebze Institute of Technology. It was supported by TUBITAK Career Project 105E097.

References

- Akgul, Y.S., Kambhamettu, C., 1999. Recovery and tracking of continuous 3d surfaces from stereo data using a deformable dual-mesh. In: Internat. Conf. on Computer Vision, pp. 765–772.
- Alvarez, L., Deriche, R., Sanchez, J., Weickert, J., 2002. Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach. *J. Visual Comm. Image Representation* 13 (1/2), 3–21.
- Aydin, T., Akgul, Y., 2006. 3D structure recovery from stereo using synchronous optimization processes. In: BMVC06, p. III:1179.
- Ben-Ari, R., Sochen, N., 2007. Variational stereo vision with sharp discontinuities and occlusion handling. In: IEEE 11th Internat. Conf. on Computer Vision, ICCV 2007, pp. 1–7.
- Ben-Ari, R., Sochen, N., 2010. Stereo matching with Mumford-shah regularization and occlusion handling. *IEEE Trans. Pattern Anal. Machine Intell.* 99 (PrePrints).
- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Machine Intell.* 23 (11), 1222–1239.
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.* 24 (5), 603–619.
- Gong, M., Yang, Y.-H., 2005. Fast unambiguous stereo matching using reliability-based dynamic programming. *IEEE Trans. Pattern Anal. Machine Intell.* 27 (6).
- Intille, S.S., Bobick, A.F., 1994. Disparity-space images and large occlusion stereo. *ECCV '94: Proc. Third European Conf. on Computer Vision*, vol. II. Springer-Verlag New York, Inc., Secaucus, NJ, USA, pp. 179–186.
- Ishikawa, H., 2003. Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Machine Intell.* 25 (10), 1333–1336.
- Kanade, T., Okutomi, M., 1994. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Trans. Pattern Anal. Machine Intell.* 16 (9), 920–932.
- Kim, H., Choe, Y., Sohn, K., 2004. Disparity estimation using a region-dividing technique and energy-based regularization. *Opt. Eng.* 43 (8), 1882–1890.
- Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Machine Intell.* 26 (2), 147–159.
- Li, G., Zucker, S.W., 2006. Differential geometric consistency extends stereo to curved surfaces. In: Proc. ECCV. Springer, pp. 44–57.
- Min, D.B., Yoon, S., Sohn, K., 2006. Segment-based stereo matching using energy-based regularization. In: MRCS, pp. 761–768.
- Ohta, Y., Kanade, T., 1985. Stereo by two-level dynamic programming. In: Internat. Joint Conf. on Artificial Intelligence, pp. 1120–1126.
- Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell. PAMI-12* (7), 629–639.
- Pock, T., Zach, C., Bischof, H., 2007. Mumford-shah meets stereo: Integration of weak depth hypotheses. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR '07, pp. 1–8.
- Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D., 2008. A convex formulation of continuous multi-label problems. In: European Conf. on Computer Vision (ECCV).
- Robert, L., Deriche, R., 1996. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. *ECCV '96: Proc. 4th European Conf. on Computer Vision*, vol. 1. Springer-Verlag, London, UK.
- Robert, L., Deriche, R., Faugeras, O.D., 1992. Dense depth recovery from stereo images. In: *ECAI '92: Proc. 10th European Conf. on Artificial Intelligence*. John Wiley & Sons, Inc., New York, NY, USA, pp. 821–823.
- Roy, S., Cox, I.J., 1998. A maximum-flow formulation of the n-camera stereo correspondence problem. In: *ICCV*, pp. 492–502.
- Saito, H., Mori, M., 1995. Application of genetic algorithms to stereo matching of images. *Pattern Recognition Lett.* 16 (8), 815–821.
- Scharstein, D., Szeliski, R., 1998. Stereo matching with nonlinear diffusion. *Internat. J. Comput. Vision* 28 (2), 155–174.
- Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Internat. J. Computer Vision* 47 (1–3), 7–42.
- Scharstein, D., Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. In: *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. I-195–I-202.
- Shah, J., 1993. A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '93*, pp. 34–40.
- Slesareva, N., Bruhn, A., Weickert, J., 2005. Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. In: *DAGM-Symposium*, pp. 33–40.
- Sun, J., Zheng, N.-N., Shum, H.-Y., 2003. Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Machine Intell.* 25 (7), 787–800.
- Tikhonov, A.N., Arsenin, V.Y., 1977. *Solutions of Ill-posed Problems*. V.H. Winston & Sons, John Wiley & Sons, Washington DC, New York.
- Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images. In: *ICCV '98: Proc. Sixth Internat. Conf. on Computer Vision*. IEEE Computer Society, Washington, DC, USA, p. 839, p. 839.

- Woodford, O., Torr, P., Reid, I., Fitzgibbon, A., 2008. Global stereo reconstruction under second order smoothness priors. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8.
- Yoon, K.-J., Kweon, I.S., 2006. Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Machine Intell. 28 (4), 650–656.
- Yang, Q., Yang, R., Davis, J., Nister, D., 2007. Spatial-depth super resolution for range images. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR '07, pp. 1–8.